2018 7th International Conference on Social Science, Education and Humanities Research (SSEHR 2018)

Graphical Statistics of Stock Market

Qi Xiao

East China Normal University, China Shanghai, 200241, China

Keywords: Clustering Coefficient; Average Shortest Path; Network Synchronization

Abstract: To analyze the statistical features of Chinese stock market, we choose 260 stocks and process the initial information by just taking trend information. Set up 3 kinds of networks and corresponding time series by using sliding window. Then model the network statistics. We conclude that the tendency of SSE Composite Index has negative effect on clustering coefficient and positive effect on average shortest path. Furthermore, average shortest path has negative effect on synchronization, and clustering coefficient has negative effect on network synchronization.

Introduction

In nature, a large number of complex systems can be described by complex networks. In the network, the vertices are elements in the system, and 2 vertices are connected by an edge. After some innovation researches of small-world networks and scale-free networks, network model is applied in many complex systems such as WWW, the Internet, the movie actor collaboration networks, cell networks and telephone networks etc. These networks have 2 common topology character of being small-world and scale-free, which were studied by some innovative researchers. Traditional network models such as regular and stochastic networks cannot embody these characters, however. Thus, researchers have built some new networks such as small-world network models and dynamic evolvement scale-free models, which make it more easily to study these characters.

In essence, stock market is a complex system. The price fluctuations among vast stocks have complicated relationships rather than changing independently. Relative researches based on network models have been proposed for studying the correlation of stock prices. Mantegna[1] was the first one to construct networks based on stock price correlations. He found a hierarchical arrangement of stocks in the stock correlation network which was built by investigating the daily time series of the logarithm of the stock price. Boginski et al.[2] analyzed the statistical features of stock networks by using threshold method and found the power-law degree distributions. Jung et al.[3], Garas et al.[4], Huang et al.[5], Tabak et al.[6], Cheong et al.[7] established networks with stocks in Korea, Athens, Brazil, Japan, and China respectively and analyzed their topological features. Lan et al.[8] did a search about all coal and electricity stocks and studied the log return in the latest 19 years.

The study on network statistics can help us understand correlation patterns among stocks, thus it can be a good warrant or guide for risk management of stock investment. Nowadays, a considerable number of stocks are traded in the Shanghai and Shenzhen stock market in China. In accordance to the data of these stocks, we construct some networks to analyze Chinese stock market.

Previous similar researches based on network usually abstract the log return as the data, but there is another approach that can refine the information of stocks. In this paper, we try to just use the tendency information about the stocks, which is more simplified than the previous way to get information. As far as closing price concerned, if the closing price of today is higher than that of yesterday, we will denote that the trend of today is 1, it is -1 if the closing price of today is lower than that of yesterday. Also, when we confront a condition that the closing price of today is equal to that of yesterday, we can denote that the trend is equal to 0. By doing this instead of calculating the log returns, we can reduce considerable amount of workload.

After processing the data, we construct 3 different kinds of networks, and define corresponding statistics that can embody topological features of each network. Then use sliding window approach to establish time series. Finally we analyze time series correlations and draw conclusions from the results.

Copyright © (2018) Francis Academic Press, UK DOI: 10.25236/ssehr.2018.054

The rest of this article is organized as follows. In section 1 we describe the processes of network construction and their topology statistical quantities of the network. Section 2 is the empirical study and results. In the last section we present some conclusions.

Construction of the Stock Network

Choose N different stocks and a period of time T. assume that the closing price of the ith stock is $P_i(t)$. Now we try to use the trend function $f_i(t)$ to define the time series we want to analyze in the following essay. $f_i(t)$ is defined below.

$$f_i(t) = \begin{cases} 1, & f_i(t+1) > f_i(t) \\ 0, & f_i(t+1) = f_i(t) \\ -1, & f_i(t+1) < f_i(t) \end{cases}, \quad t \in [1, T-1]$$

in this way, a time series $f_i(t)$ is set up.

Using Pearson correlation clustering to measure the correlation between 2 trends of stocks $f_i(t)$ and $f_j(t)$, the correlation of stock i and j is defined as:

$$C_{ij} = \frac{E(f_i f_j) - E(f_i)E(f_j)}{\sqrt{Var(f_i)Var(f_j)}}$$

$E(f_i)$ is the average of all $f_i(t)$ in T days, while $Var(f_i)$ is the variance of all the $f_i(t)$ in T days.

$$E(f_i) = \frac{1}{T}\sum_{t=1}^{T} f_i(t)$$

$$Var(f_i) = \frac{1}{T}\sum_{t=1}^{T}\left(f_i(t) - E(f_i)\right)^2$$

Recognize the N stocks as N notes in the network, and C_{ij} is regarded as the weight of edge between ith and jth.

In this paper, we try to establish 3 different kinds of network models.

A. Threshold Network

When establishing the threshold network, we should decide the threshold $\theta(\theta\square[0,1])$. After deciding θ, we can define the edge set W_θ:

$$W_\theta = \begin{cases} w_{ij} = 1, C_{ij} > \theta, \\ w_{ij} = 0, C_{ij} \le \theta. \end{cases}$$

The network has N vertices and the weight of edge between i and j is w_{ij}. We denote this network as G_θ.

B. Distance Network

According to Mantegna[9] theory, correlation coefficient C_{ij} can be converted into distance between stock i and j. naturally, distance can be defined with the following equation.

$$D_{ij} = \sqrt{2 * \left(1 - C_{ij}\right)}$$

so we can define G_D as a network with N vertices and the edge between i and j has weight D_{ij}.

C. Weight Network

Using Peron and Rodrigues'[10] approach to establish stock network, we can convert distance network into another kind of network named "weight network". The edge weight between i and j is defined as below:

$$\omega_{ij} = \exp(-D_{ij})$$

The network has N vertices and the weight of edge between i and j is ω_{ij}. We denote this network as G_ω.

Statistics

In this paper, we mainly take small-world property and non-scale property into consideration. Small-world[11] property means that there is strong correlation between stocks. Clustering coefficient, average shortest path and network synchronization will be analyzed in this paper because Watts held that small-world property should defined as a network which has relatively large clustering coefficient and small average shortest path.

A. Clustering Coefficient

Clustering coefficient is used to describe the clustering features of a network. When it comes to the stock network, clustering coefficient describes the N stocks gather. Assume that vertex i is connected with other k_i vertices. There are no more than $k_i(k_i-1)/2$ edges between these k_i vertices, while there are E_i edges. CC_i could be defined.

$$CC_i = \frac{2E_i}{k_i(k_i-1)}$$

and the clustering coefficient of the whole network is defined as

$$CC = \frac{\sum_{i=1}^{N} CC_i}{N}$$

B. Average Shortest Path

Average shortest path describes the speed of information transformation between 2 stocks. If the average shortest path is shorter, the transformation speed of the whole network will be more swift.

If i and j are connected by some edges, shortest path length is the minimum addition of all the edge weights. The average shortest path length is defined as:

$$L = \frac{1}{\frac{1}{2}N(N-1)} \sum_{i \le j} SD_{ij}$$

SD_{ij} is the shortest path connecting vertex i and j, and it is derived from Dijkstra[12] algorithm.

C. Synchronization

Synchronization is a conception that can describe how the price of stocks go up or go down. If stocks have strong synchronization, they incline to change in the same direction (they may all go up, or they may all go down). Otherwise they might change in the different directions. According to Liu and Tse's[13] definition of network synchronization, in this paper, synchronization of stock network can be defined as below:

$$M = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1, j \ne i}^{N} \omega_{ij}$$

ω_{ij} is the edge weight between vertex i and j in the definition of G_ω mentioned above.

In this paper, time series are established through sliding window method. Assume that the window length is r, and the time span is T, same as the stock data we've got. Firstly, choose the 1st day and r+1th day as the window, and then establish 3 kinds of networks described above. Calculating clustering coefficient, average shortest path length and network synchronization is the last stage. By imitating the calculation process above, we can also calculate 3 statistics with window 2nd to the r+1th day. In this way, we can deduce the rest of the time series. Every kind of statistic time series is a (T-r-1)-length series.

Empirical Study and Results

To analyze Chinese stock market, we found 260 stocks and calculate 3 different statistics mentioned above.

A. Clustering Coefficient

let θ range from 0.1 to 0.9, we can get 9 different curves named $CC_{0.1}$~$CC_{0.9}$. There is a picture showing their changes.

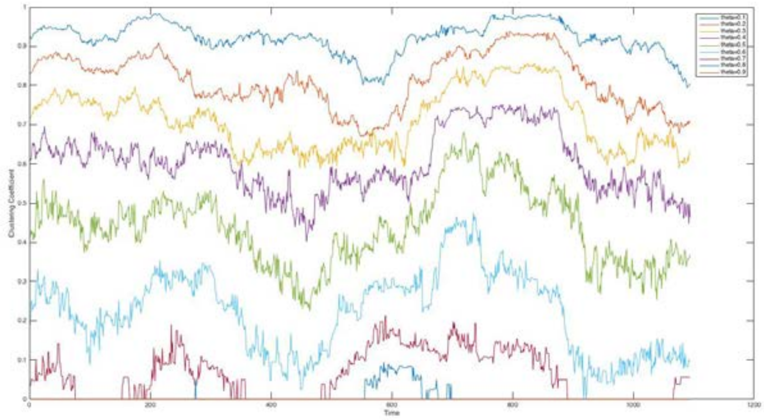

Figure 1 Time series of clustering coefficient under different threshold

Table 1 Regression coefficients of clustering coefficient under different threshold

Coefficient	$CC_{0.1}$	$CC_{0.2}$	$CC_{0.3}$	$CC_{0.4}$	$CC_{0.5}$	$CC_{0.6}$	$CC_{0.7}$	$CC_{0.8}$	$CC_{0.9}$
$CC_{0.1}$	1	1.156	—	—	—	—	—	—	—
$CC_{0.2}$	0.917	1	0.930	—	—	—	—	—	—
$CC_{0.3}$	—	0.990	1	0.919	—	—	—	—	—
$CC_{0.4}$	—	—	0.876	1	1.072	—	—	—	—
$CC_{0.5}$	—	—	—	0.918	1	0.912	—	—	—
$CC_{0.6}$	—	—	—	—	0.974	1	0.461	—	—
$CC_{0.7}$	—	—	—	—	—	1.045	1	0.162	—
$CC_{0.8}$	—	—	—	—	—	—	1.51	1	0
$CC_{0.9}$	—	—	—	—	—	—	—	NA	1

Although clustering coefficient time series are totally different under different thresholds, it is obvious that the curves have similar tendency. We can deduce that the stock market has the same feature under different thresholds from the coefficients in Table 1. Since the regression coefficients under thresholds next to a certain threshold are around 1, it could be explained easily why curves under different thresholds have similar tendency. As we all know, correlation is significant when correlation is larger than 0.5. in the following content we use $CC_{0.5}$ to represent the clustering coefficient series of the whole stock market and denote it as CC.

B. Average Shortest Path and Synchronization

Calculating the shortest path length and synchronization, we can get 2 series L and M. Draw their time series before analyzing.

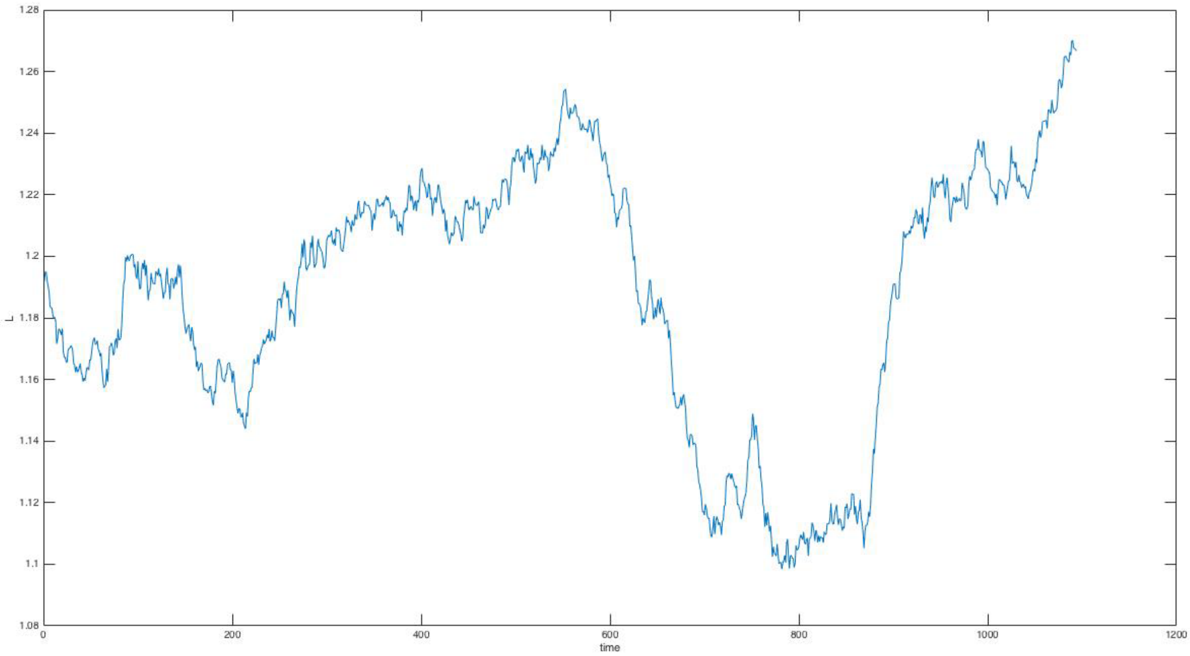

Figure 2 Time series of average shortest path

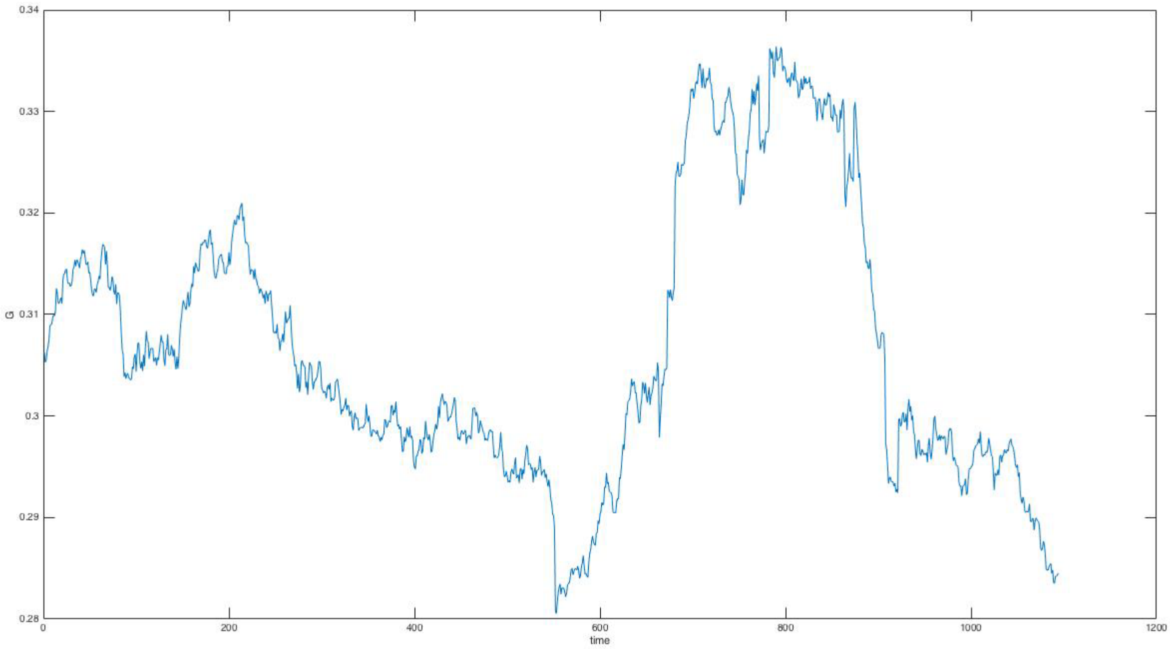

Figure 3 Time series of synchronization

From Figure 2 and Figure 3 we can infer that for one specific period of time, the shortest average is relative small while the corresponding synchronization is strong. It is easily to explain: the shorter the path between stocks are, the faster the information will transfer. Then the stocks tend to change in the same direction, which means that the synchronization will be strengthened.

Table 2 Statistics of clustering coefficient and average shortest path

series	maximum	minimum	average	variance
L	1.2701	1.0984	1.1880	0.0018
CC	0.6814	0.2246	0.4385	0.0091

From Table 2 we can infer that the stock network has relatively small average shortest path length, while CC is large. That means the stock network holds small-world property.

C. Cointegration Regression of Series

Choose 260 stocks in china stock market and analyze those statistics. We also choose SSE (Shanghai Stock Exchange) Composite Index and use trend function to describe its fluctuation. $I(t)$ is the closing point of the SSE Composite Index at t day. $S(t)$ is the average trend of the SSE Composite Index. Define:

$$S(t) = \frac{1}{r+1}\sum_{k=t}^{t+r} f(k)$$

Where $f(t) = \begin{cases} 1, f(t+1) > f(t) \\ 0, f(t+1) = f(t) \\ -1, f(t+1) < f(t) \end{cases}$, $t \in [1, T-1]$.

In the rest of this section, we analyze these time series by cointegration regression method. Above all, we try to analyze stationarity of each time series. We do the ADF test and find out that the 4 series are all not stationary. In this case we calculate the difference time series of each time series and do the ADF test. It is obvious that the first difference time series are all stationary. Therefore, $CC, S, L, M \sim I(1)$,which means that they meet the cointegration requirement. Under this condition, we can use cointegration regression to set up linear models. Next, we estimate corresponding coefficients using OLS (Ordinary Least Square) method. Then we run stationarity test on the estimated residual series to confirm the accuracy of models.

$$CC = \alpha_0 + \alpha_1 S + \varepsilon \tag{1}$$

$$L = \beta_0 + \beta_1 S + \varepsilon \tag{2}$$

Table 3 Estimate of coefficient in regression model (1)

coefficient	α_0	α_1
estimation	0.460117	-0.254442

Table 4 Performance in regression model (1)

Value of test-statistic	-3.1303		
Critical values for test statistics:			
	1pct	5pct	10pct
tau1	-2.58	-1.95	-1.62

From Table 3 and Table 4, we can infer that S has negative effect on CC. And it accounts for 16.9% proportion of CC.

Table 5 Estimate of coefficient in regression model (2)

coefficient	β_0	β_1
estimation	1.175929	0.142441

Table 6 Performance in regression model (2)

Value of test-statistic	-2.8206		
Critical values for test statistics:			
	1pct	5pct	10pct
tau1	-2.58	-1.95	-1.62

Table 5 shows that S has positive effect on L, while Table 6 indicate that S account for around 11% of L.

In a conclusion, when the whole market tends to decline, the correlation between stocks gets tighter, which can be infer from larger CC and smaller L.

To analyze the topological features of stock market, we try to analyze the relation between G and L. Build 2 other models:

$$M = \gamma_0 + \gamma_1 CC + \varepsilon \tag{3}$$

$$M = \delta_0 + \delta_1 L + \varepsilon \tag{4}$$

Table 7 Estimate of coefficient in regression model (3)

coefficient	γ_0	γ_1
estimation	0.256839	0.113628

Table 8 Performance in regression model (3)

Value of test-statistic	-3.935		
Critical values for test statistics:			
	1pct	5pct	10pct
tau1	-2.58	-1.95	-1.62

From Table 7 and Table 8 we can see that CC has a positive effect on M, which means that the larger CC is, the stronger synchronization will be.

Table 9 Estimate of coefficient in regression model (4)

coefficient	δ_0	δ_1
estimation	0.684697	-0.318201

Table 10 Performance in regression model (4)

Value of test-statistic	-4.5894		
Critical values for test statistics:			
	1pct	5pct	10pct
tau1	-2.58	-1.95	-1.62

What we can infer from these 2 tables is that the L has negative effect on M. This means that if the average shortest path length is small, the synchronization will be strong because of swift information transformation speed.

To preclude chanciness, window length is changed in 3 different lengths, which is set as 30,60,90,120 days. Same conclusion can be inferred despite of different window lengths, which indicates that the conclusion we draw from experiments is independent on window length.

Conclusions

Complex networks have been constructed for Chinese stock market (choose 260 stocks in traded in Shanghai and Shenzhen stock market, from 2012-06-01 to 2017-06-01). In the first stage we process the closing price and turn it into tendency information represented by 1, -1, and 0. Then we construct 3 different kinds of networks and calculate corresponding statistics. Furthermore, we use sliding window method to set up 3 time series of statistics and find the relationships between these time series, using cointegration regression method.

The result suggests that different threshold does not influence the tendency of stocks. Another conclusion we can draw from the theory and empirical study is that the topological features can reflect the tendency of the whole stock market. When the synchronization rises, stock prices tend to change in the same direction. At the same time, the relation between stocks is becoming tighter (smaller average shortest path and larger clustering coefficient). Then the whole stock market will decline with great possibility. On the contrary, if the synchronization falls in a period of time, the stock prices change in different directions, so these stocks are not so closely related. In that case, the whole stock market may have a positive tendency, which is a good news to investigators. It is meaningful to use network when analyzing stock market.

References

[1].Bonanno, G., F. Lillo, and R. N. Mantegna. 'High-Frequency Cross-Correlation in a Set of Stocks', Quantitative Finance, vol. 1/no. 1, (2001), pp. 96-104.

[2]. Boginski, Vladimir, Sergiy Butenko, and Panos M. Pardalos. 'Statistical Analysis of Financial Networks', Computational Statistics and Data Analysis, vol. 48/no. 2, (2005), pp. 431-443.

[3]. Jung, Woo-Sung, Seungbyung Chae, Jae-Suk Yang, et al. 'Characteristics of the Korean Stock Market Correlations', Physica A: Statistical Mechanics and its Applications, vol. 361/no. 1, (2006), pp. 263-271.

[4]. Garas, Antonios, and Panos Argyrakis. 'Correlation Study of the Athens Stock Exchange', Physica A: Statistical Mechanics and its Applications, vol. 380/(2007), pp. 399-410.

[5]. HUANG Feixue, GU Jing, Li Yanxi, et al. Before and after the financial crisis of the world's main stock index linkage compared with dynamic stability [J] .Systems Engineering Theory and Practice, 2010,30(10),pp.1729-1740.

[6]. Tabak, Benjamin M., Thiago R. Serra, and Daniel O. Cajueiro. 'Topological Properties of Stock Market Networks: The Case of Brazil', Physica A: Statistical Mechanics and its Applications, vol. 389/no. 16, (2010), pp. 3240-3249.

[7]. Cheong, SA, RP Fornia, GHT Lee, et al. 'The Japanese Economy in Crises: A Time Series Segmentation Study', Economics-the Open Access Open-Assessment e-Journal, vol. 6/no. 2012-5, (2012), pp. 1.

[8]. LAN Wangsen, ZHAO Guohao. Application of complex network to study the strong correlation of stock in plate [J] .Journal of Sun Yat-Sen University (Natural Science Edition), 2010, 49(sl):65-69.

[9]. Mantegna R N, Stanley H E, Chriss N A. An Introduction to Econophysics: Correlations and Complexity in Finance[J]. Physics Today, 2000, 53(12):70-70.

[10]. Peron T K D, Rodrigues F A. Collective behavior in financial market[J]. Quantitative Finance, 2011, 96(4):48004.

[11]. Borsboom, D., A. O. J. Cramer, V. D. Schmittmann, et al. 'The Small World of Psychopathology', PLoS One, vol. 6/no. 11, (2011), pp. e27407.

[12]. Sanghani H. dijkstra algorithm[J]. Communications Networks.

[13]. Tse, Chi K., Jing Liu, and Francis C. M. Lau. 'A Network Perspective of the Stock Market', Journal of Empirical Finance, vol. 17/no. 4, (2010), pp. 659-667.